ネットワーク侵入検知システムの高度化に関する研究

氏名（本籍地）	山田（やまだ） 明（あきら）
学位の種類	博　士（情報科学）
学位記番号	情　博　第　456　号
学位授与年月日	平成21年　3月25日
学位授与の要件	学位規則第4条第1項該当
研究科、専攻	東北大学大学院情報科学研究科（博士課程）応用情報科学専攻
学位論文題目	ネットワーク侵入検知システムの高度化に関する研究
論文審査委員	（主査）東北大学教授　加藤　寧
	東北大学教授　橋本　和夫　　東北大学教授　篠原　歩

論文内容の要旨

1章 序論

あらゆるサービスが電子化されネットワークを介して提供される中、サービスを対する不正アクセスが深刻な問題となっている。そして、サービスへの不正アクセスを監視する侵入検知システム（IDS :Intrusion Detection System）の重要性が高まっている。

侵入検知システムは、ネットワーク回線やサービス提供ホストを監視して、ネットワークやホストへの不正な侵入やアクセス制御ポリシの侵害を検知するシステムである。現在、侵入検知システムは多くの方法が提案され製品化されている。しかしながら、侵入検知システムには未だに解決されていない課題が残っている。まず、不正があらかじめ定義されていない未知の攻撃を侵入検知システムが検知することは困難である。次に、監視対象の通信が暗号化されている場合にネットワーク型 侵入検知システムによる監視が困難である。さらに、侵入検知システムは誤検知を含む大量の検知結果を出力するため、管理者による解析が困難である。

本論文では、ネットワークを流れる通信を監視してサービスに対する不正なアクセスを検知するネットワーク型 侵入検知システムにおいて、未知攻撃の課題、暗号化された通信の課題、検知結果の解析における課題に対する解決策を示す。

2章 教師情報を自動生成する機械学習によるアノマリ検知

侵入検知システムの検知方法は、ミスユース検知とアノマリ検知に分類され、それぞれに利点と欠点が存在する。まず、ミスユース検知は、ミスユースをあらかじめ定義するため誤検知が少ないが、ミスユースとして定義されていない未知攻撃を検知できない。一方、アノマリ検知は、学習データに存在しない未知攻撃であっても通常状態との違いから検知できるが、比較的多くの誤検知を発生する。したがって、侵入検知システムは、未知攻撃を高い精度で検知することが困難である。

本章では、ミスユース検知とアノマリ検知を組み合わせることによって、未知攻撃の課題を解決する方法を提案する。侵入検知システムは、ミスユース検知において未知攻撃を検知することが全くできず、アノマリ検知において誤検知を多く発生してしまう。特に、機械学習によるアノマリ検知は、教師あり学習の方が誤検知の発生が少ない。しかし、教師あり学習のためには、教師情報を用意する必要がある。教師情報は、専門家が学習データを解析して、攻撃の種類や発生時刻を抽出したものである。しかし、専門家による解析が必要であるため、時間的および金銭的な費用が発生する。

そこで、ミスユース検知の検知結果を利用することによって、教師情報を自動的に生成する。ミス

ユース検知には、攻撃の種類や発生時刻が含まれているため教師情報として利用できる。また、ミスユース検知は、多数の製品が存在するため、ミスユースの定義が定期的に追加更新される。そのため、更新されたミスユースによって、最新の教師情報が生成できる。提案方法の概念を検証するために、HTTP に特化したアノマリ検知による侵入検知システムを実装して、既存のミスユース検知による侵入検知システムと組み合わせて評価する。

評価には、侵入検知システム評価用の公開データセットと実ネットワークから収集したデータセット、そして脆弱性監査ツールによって発生させた疑似攻撃を含むデータセットの 3 種類を用いた。まず、それらのデータセットにおいて、ミスユース検知の検知結果をから、機械学習に利用できる教師情報を自動的に生成できること検証した。

また、ミスユース検知によって生成した教師情報を利用して学習データに含まれない未知攻撃を検知できることを確認した。侵入検知システム評価用データセットでは、学習データに攻撃が含まれている場合すべての攻撃を検知でき、含まれていない場合に 7 種類の攻撃のなかで 3 種類を検知できた。また、実ネットワークから収集したデータセットと脆弱性監査ツールによるデータセットでは、検知率 67%、誤検知率 0.005%と検知率 82%、誤検知率 0%を達成した。

3 章 暗号化された通信に対する侵入検知

侵入検知システムの監視方法はホスト型とネットワーク型に分類される。ホスト型 侵入検知システムは、ホストにインストールする必要があるため、サービス提供ホストの計算機資源を消費してしまい、サービス自身に影響を与える可能性がある。一方、ネットワーク型 侵入検知システムは、ネットワークの通信のみを監視するため、監視に使用できる情報が制限される。特に、通信がホストと端末間で暗号化されている場合、通信から得られる情報が著しく制限され、攻撃を検知することが困難である。したがって、侵入検知システムは、暗号化された通信における攻撃の検知が困難である。

本章では、暗号化された通信を復号せずに、含まれている攻撃を検知する方法を示す。SSL/TLS のような暗号化プロトコルは、効率的に通信するために、全ての情報の機密性を確保するように設計されていない。例えば、送受信されるデータは、あるデータサイズに分割されて個別に暗号化されるため、通信の観測によってそのサイズや送受信時刻を抽出できる。つまり、データサイズ、通信方向、タイミングから暗号化されている通信の内容を推測できる。

具体的には、SSL/TLS 通信における Web サーバへのリクエストやレスポンスのサイズを推測できる。さらに、Web ページを構成するファイルサイズの組合せや、ユーザによるページ閲覧順序とタイミングを推測できる。暗号化されていない通信における侵入検知に利用される特徴のなかで、推測した通信内容から利用できる特徴のみを組み合わせて攻撃を検知する。

評価には、侵入検知システム評価用の公開データセットと実ネットワークから収集したデータセットを SSL/TLS による暗号化と同等の処理を行って利用する。つまり、データのペイロードに関する情報を削除し、パケットごとにプロトコルで規定されているパディング処理をした。

それらのデータセットにおいて、暗号化された通信であっても攻撃を提案方法が検知できることを確認した。また、侵入検知システム評価用データセットでは、暗号化されているにも関わらず、検知率 75%誤検知率、25%を達成した。実ネットワークから収集したデータセットでは、検知率 95%、誤検知率 5%を達成した。これらの検知率は侵入検知システムとして高くないものの、暗号化された通信に対する検知として高い値である。

4章 要約による検知結果の解析

侵入検知システムを有効に利用するためには、システムを設置するだけでなく運用する必要がある。通常の侵入検知システムは、ある程度の誤検知率を許容して実装されるため、誤検知を含む検知結果を大量に出力する。また、複数の侵入検知システムが設置される場合や、ルータやファイアウォールなどの他のネットワーク機器が存在する場合、複数の検知結果やログを解析する必要がある。

本章では、侵入検知システムの検知結果に限らず、様々なネットワーク機器のログを対象にする。ここで、侵入検知システムの検知結果やネットワーク機器のログは、膨大な量であり連続的に出力されるという特徴がある。連続的で大量のログを要約ことによって解析する方法を提案する。

ログ要約のために、データベースを要約するアルゴリズムの1つである属性指向帰納を、連続的なログや多様なフォーマットに対応できるように改良する。特に、属性指向帰納が必要とする概念階層と呼ばれる階層構造を、ログの頻度から適応的に構成する方法を提案する。具体的には、ログの頻度に基づいて階層的クラスタリングすることによって、概念階層として利用できる階層構造を構成する。さらに、単位時間ごとに分割したログに対して階層的クラスタリングすることによって、適応的に概念階層を構成できる。提案方法を用いることによって、例えば IP アドレスにおける概念階層を適応的に構成でき、その概念階層を利用して属性指向帰納を実行できる。

提案方法の有効性を検証するために、侵入検知システムの検知結果とネットワーク機器の1つである FTP サーバログとを用いて評価する。評価に使用する FTP サーバは、Linux のディストリビューションなどの多数 FTP サーバのミラーになっており、国内外からの大量のアクセスがあるため 1 日に 20 万行のログを出力する。また、侵入検知システムは、その FTP サーバの回線を監視しているため、1 日に約 15 万行の検知結果を出力する。

この評価の結果、侵入検知システムを含む多様なフォーマットに対して提案方法を適用できることが分かった。そして、連続的に出力されるログに対して、属性指向帰納に適用できる概念階層が適応的に構成できることを確かめた。また、Intel Xeon 2.6GHz の CPU で 10 万行のログを約 700 秒で解析できることが分かった。さらに、実際のログに含まれる異常を検知できることを定性的に確認した。

5章 結論

ネットワークよって提供されるサービスに対する不正アクセスを監視するために侵入検知システムが提案されている。未知攻撃の課題、暗号化された通信の課題、検知結果の解析の課題は、侵入検知システムが提案された当初から指摘されている課題であるが未だに根本的な解決策は存在しない。これまでに、さまざまな解決策が提案されてきたが、技術の発展やサービスの多様化に伴い、新しい攻撃が発生してきている。しかしながら、全体として、侵入検知システムに代表とされるセキュリティ技術の適用範囲が拡大する傾向にあり今後も継続すると思われる。本論文は、未知攻撃、暗号化された通信、検知結果の課題に対して、それぞれひとつの解決策を示すものである。これらの、解決策は、より広い範囲にセキュリティ技術を適用するということ技術の発展の中で、ネットワーク侵入検知システムの高度化に寄与できたといえる。

論文審査結果の要旨

コンピュータネットワークの発達に伴い、あらゆるサービスの電子化やネットワーク化が急速な勢いで進んでいる。昨今ネットワークサービスに対する不正アクセスが問題となっている。ネットワーク上のサービスを監視するためにネットワーク侵入検知システムが有効であるが、サービスの多様化、攻撃方法の巧妙化に伴って新たな課題に直面している。著者は、監視対象となるサービスの拡大や検知対象とする攻撃の複雑化に対応すべく侵入検知システムの高度化に関する研究を行ってきた。本論文は、その成果をまとめたもので全編5章からなる。

第1章は序論である。侵入検知システムの歴史と分類に始まり、侵入検知システムにおける現状の課題を未知攻撃の課題、暗号化された通信の課題、検知結果の解析における課題としてまとめている。侵入検知システムの背景を体系的に整理しており、その中で本研究の位置づけを明確化している点において評価できる。

第2章では、未知攻撃の課題に対する解決策として教師情報を自動生成するアノマリ検知による侵入検知システムを提案している。これは、過去に観測されたことのない未知攻撃を検知できるアノマリ検知において、検知精度の向上につながる学習データと教師情報をミスユース検知により自動的に生成する技術である。実ネットワークのデータセットを用いる実験により有効な教師情報を生成できることを確認している。これは、現在でも課題となっている未知攻撃検知の課題解決に大きく貢献したものである。

第3章では、暗号化された通信における侵入検知を提案している。これは、従来の侵入検知システムが検知できなかった暗号化された通信における攻撃を、データサイズやタイミング情報を利用することによって通信を復号することなく検知する技術である。実ネットワークにおけるデータセットを用いる実験により暗号化された攻撃を検知できることを確認している。暗号化された通信を対象とする侵入検知システムは新しい試みであり高く評価できる。

第4章では、検知結果の解析における課題に対して、検知結果を要約することにより重要な情報を抽出する方法を提案している。データベースを要約するアルゴリズムである属性指向帰納を侵入検知システムの検知結果に適用するにあたり必要となる概念階層を検知結果の頻度に合わせて適応的に生成する方法が述べられおり、侵入検知システムを含む様々なネットワークログに解析対象を拡大できる。これは、膨大な検知結果から重要な情報を抽出することを可能にし、検知結果の解析効率を大幅に向上する技術を創出したものであり高く評価できる。

第5章は、結論である。

以上要するに本論文は、ネットワーク侵入検知システムにおいて、検知する攻撃の対象や監視するネットワーク対象の拡大を実現するものであり、情報通信技術ならびに応用情報科学の発展に寄与するところが少なくない。

よって、本論文は博士（情報科学）の学位論文として合格と認める。